

基于改进共同因子模型的死亡率预测模型

李乐乐, 肖鸿民

西北师范大学数学与统计学院, 甘肃 兰州

收稿日期: 2023年5月21日; 录用日期: 2023年6月14日; 发布日期: 2023年6月21日

摘要

人口死亡率的研究是与长寿风险相关的金融衍生品开发的理论依据, 其研究的关键在于建立适合我国国情的死亡率模型, 提高死亡率预测的准确性, 考虑到中国人口死亡率数据较少且质量存在问题, 仅使用我国死亡率数据建模不易得到好的预测结果。本文对共同因子模型进行改进创新, 在其本身结构上加上经济项, 并使用三个发展中国家和三个发达国家的历史死亡率数据进行拟合预测, 结果表明, 本文提出的新模型效果较好, 弥补了仅依赖人口死亡率建模的缺点。

关键词

人口死亡率预测, 多总体随机死亡率模型, 人均GDP, 最小二乘法

Mortality Prediction Model Based on Improved Common Factor Model

Lele Li, Hongmin Xiao

School of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: May 21st, 2023; accepted: Jun. 14th, 2023; published: Jun. 21st, 2023

Abstract

The study of population mortality is the theoretical basis for the development of financial derivatives related to longevity risk, and the key to its research is to establish a mortality model suitable for China's national conditions and improve the accuracy of mortality prediction. Considering the small amount and quality of China's population mortality data, it is not easy to get good prediction results by modeling using only China's mortality data. This paper improves and innovates the common factor model by adding an economic term to its own structure and fitting the prediction using historical mortality data from three developing countries and three developed countries, and the results show that the new model proposed in this paper works better and makes up for

the shortcomings of relying only on population mortality modeling.

Keywords

Population Mortality Projections, Multi-Total Stochastic Mortality Model, GDP Per Capita, Least Squares

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人口死亡率是寿险精算产品和基本养老金定价的基础,也是与长寿风险相关的金融衍生品开发的理论依据。利用随机动态死亡率模型进行死亡率预测是业界和学术界都十分关注的问题。目前,对多个人口的联合动态预测是研究热点,因为即使人们只对某一特定国家的预测感兴趣,多人口模型也有合并更多数据的优势。正如 Li 和 Lee 所指出的,通过考虑更大群体中的共同模式,可以改善单个人口的死亡率预测[1]。

目前常用的多人口随机死亡率模型主要包括 Joint-k 模型、ACF 模型及其改进模型和 CAE 模型,这些模型通常将死亡率数据分解为时间因素、年龄因素来描述死亡率变化趋势,进而依靠预测时间因素来预测死亡率,是一种仅根据历史趋势外推预测死亡率的方法[2]。但我国对死亡率数据的统计工作开展地较晚且由于人口基数过大,数据质量较差,直接应用经典随机死亡率模型并不能取得较好的效果。

与此同时,为了弥补仅依赖人口死亡率建模预测的缺点,一些外生因素(如生物学、环境或社会经济因素)对死亡率影响的研究也层出不穷,其中,研究最多的趋势之一是经济增长。例如,2005年, Brenner 指出,至少十年来的经济增长是美国二十世纪死亡率下降的主要因素[3]。2011年, Hanewald 使用 1950年至 2006年期间六个经合组织国家的数据,在 Lee Carter 死亡率预测模型中研究了宏观经济波动对死亡率的动态影响[4],发现该模型中的死亡率指数与所考虑时期内国家的 GDP 水平之间存在显著相关性。2014年, Niu 和 Melenberg 建立了死亡率指数 k_t 和宏观经济指标的向量误差校正模型(VECM) [5],并对 Lee-Carter 模型中的时间因素和人均 GDP 进行了 Johansen 协整检验,结果表明两者存在长期稳定的均衡关系,这种协整关系比 Hanewald 提出的结果强得多。Boonen 和 Li 在 2017 年对上述做法进行扩展[6],将以人均实际国内生产总值(GDP)为代表的经济增长包括在内,扩展了 Li 和 Lee 在 2005 提出的模型[1]以捕捉具有相似社会经济条件的人群的共同死亡率趋势,发现提出的模型提供了更好的样本内拟合和样本外预测性能。2022年,王梦珂和唐爽对 Boonen 和 Li 的模型加以改进[7],把作为全局参数的人均 GDP 变量 $G(t)$ 变为局部参数 $g(t,i)$,打破了只能选择经济发展趋势相近的国家作为一组进行建模的局面,研究了包括中国在内的 5 个国家,结果表明改进后的新模型适合中国人口死亡率的预测,弥补了仅依赖人口死亡率建模预测的缺点。

本文受前人研究启发,创新地在共同因子模型 ACF1, ACF3 的本身结构上加上经济项,这里采用的是标准化后的人均 GDP 值,并将新模型命名为 ACF1-GDP 和 ACF3-GDP,使用 ACF1-GDP、ACF1、ACF3-GDP、ACF3 这四个模型分别对三个发展中国家(中国、保加利亚、智利)和三个发达国家(美国、日本、英国) 1995~2017 年的历史死亡率数据进行参数拟合,并对中国、智利、美国、日本、英国 2018~2020

年, 保加利亚 2018~2021 年的死亡率进行预测, 最后将 ACF1-GDP、ACF1、ACF3-GDP、ACF3 预测结果误差用绝对拟合误差 AFE 来度量, 以此探究加上经济项后的效果是否良好, 结果表明提出的新模型 ACF1-GDP 和 ACF3-GDP 的预测误差对所研究的大部分国家在大多数时候要比 ACF1、ACF3 的预测误差小, 证明本文对共同因子模型的改进效果良好。

2. 模型介绍

2.1. 模型提出

Li 等[1]考虑多人口之间的共同趋势, 提出了增强共同因子模型, 即 ACF 模型:

$$\log m(x, t, i) = a(x, i) + B(x)K(t) + b(x, i)k(t, i). \tag{1}$$

Enchev 等[8]将 ACF 模型中的 $b(x, i)$ 换为 $B^{(2)}(x)$, 提出了 ACF1 模型:

$$\log m(x, t, i) = a(x, i) + B^{(1)}(x)K(t) + B^{(2)}(x)k(t, i). \tag{2}$$

马海飞[9]对 ACF 模型变形, 将其中的 $k(t, i)$ 换为 $K^{(2)}(t)$, 提出了 ACF3 模型:

$$\log m(x, t, i) = a(x, i) + B(x)K^{(1)}(t) + b(x, i)K^{(2)}(t). \tag{3}$$

对于 ACF1 模型, $m(x, t, i)$ 表示第 i 个人群中 x 岁的人在年份 t 的中心死亡率, $a(x, i)$ 描述特定年龄人口的平均死亡率水平, 且有:

$$a(x, i) = \frac{\sum_t \log m(x, t, i)}{T}. \tag{4}$$

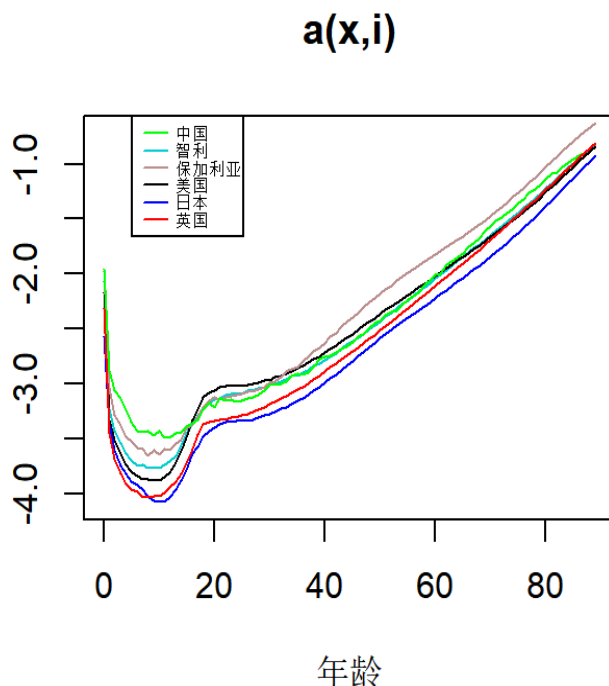


Figure 1. Average mortality levels by country ($a(x, i)$)

图 1. 各国平均死亡率水平($a(x, i)$)

从图 1 看出, 各个国家的对数死亡率的平均水平随着年龄的增长, 其值均呈现了先下降后上升的趋

势, 体现了不同国家对数死亡率的平均水平随着年龄的增长变化的一种共性。

$K(t)$ 为共同时间因子, 反映全年龄总体人口对数死亡率随时间变化的共同趋势; $k(t,i)$ 为附加时间因子, 反映第 i 个人群对数死亡率随时间变化的趋势; $B^{(1)}(x)$, $B^{(2)}(x)$ 为共同年龄因子, 分别表示特定年龄总体人口对 $K(t)$ 和 $k(t,i)$ 变化的敏感度。

对于 ACF3 模型, $K^{(1)}(t), K^{(2)}(t)$ 为共同时间因子, $B(x)$ 为共同年龄因子, $b(x,i)$ 是附加年龄因子, 代表第 i 个人群对共同时间因子 $K^{(2)}(t)$ 的敏感度。

本文分别在 ACF1 和 ACF3 模型后加上 $\gamma(x,i)$ 和 $g(t,i)$ 的乘积项, $g(t,i)$ 表示第 i 个人群在 t 年的标准化后的人均 GDP 值, 将 $g(t,i)$ 标准化后用于建模, 这意味着本文考虑的是经济变化趋势对人口死亡率变动的影响, $\gamma(x,i)$ 反映特定年龄人口死亡率对经济因素变化的敏感程度。并将新模型分别命名为 ACF1-GDP 和 ACF3-GDP, 模型形式如下:

ACF1-GDP:

$$\log m(x,t,i) = a(x,i) + B^{(1)}(x)K(t) + B^{(2)}(x)k(t,i) + \gamma(x,i)g(t,i). \quad (5)$$

ACF3-GDP:

$$\log m(x,t,i) = a(x,i) + B(x)K^{(1)}(t) + b(x,i)K^{(2)}(t) + \gamma(x,i)g(t,i). \quad (6)$$

人均GDP(标准化)

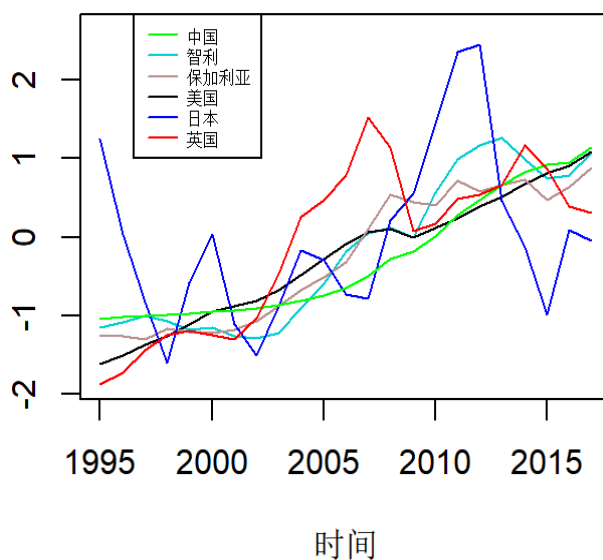


Figure 2. Average mortality levels by country ($a(x,i)$)

图 2. 标准化后人均 GDP ($g(t,i)$)

从图 2 看出, 只有中国标准化后的人均 GDP 没有明显的下降趋势, 其他国家图线走向均有明显变化, 这意味着六国中只有中国的经济水平始终在平稳增长。

2.2. 参数估计

对于 ACF1-GDP, $B^{(1)}(x)$, $B^{(2)}(x)$, $K(t)$ 为全局参数, $k(t,i)$, $\gamma(x,i)$, $g(t,i)$ 为局部参数, 且 $g(t,i)$ 为已知量。

该模型的参数估计大致通过以下两步完成:

第一步: 将所有人口死亡率用于计算全局参数;

第二步: 在全局参数已知的条件下, 用最小二乘法拟合未知的局部参数。

$$U(x,t) = \frac{\sum_i (\log m(x,t,i) - a(x,i))}{I}, \quad (7)$$

其中, I 表示样本内人口类别总数, $U(x,t)$ 包含所有人口死亡率信息, 被用于第一步全局参数的计算。通过对 $U(x,t)$ 进行奇异值分解得到全局参数 $B^{(1)}(x)$, $B^{(2)}(x)$, $K(t)$ 。然后令

$$U_i(x,t) = B^{(1)}(x)K(t) + B^{(2)}(x)k(t,i) + \gamma(x,i)g(t,i). \quad (8)$$

根据最小二乘法[10]得到以下参数迭代公式:

$$k(t,i) = \frac{\sum_{x=1}^n B^{(2)}(x) (U_i(x,t) - B^{(1)}(x)K(t) - \gamma(x,i)g(t,i))}{\sum_{x=1}^n B^{(2)}(x)^2}, \quad (9)$$

$$r(x,i) = \frac{\sum_{t=1}^T g(t,i) (U_i(x,t) - B^{(1)}(x)K(t) - B^{(2)}(x)k(t,i))}{\sum_{t=1}^T g(t,i)^2}, \quad (10)$$

基于 Niu 和 Melenberg [5]、Boonen 和 Li [6] 的做法, 为了确保参数估计结果的唯一性, 添加以下约束:

$$\sum_t k(t,i) = 0. \quad (11)$$

对于 ACF3-GDP, $K^{(1)}(t)$, $K^{(2)}(t)$, $B(x)$ 为全局参数, $b(x,i)$, $\gamma(x,i)$, $g(t,i)$ 为局部参数, 与 ACF1-GDP 拟合步骤类似, 局部参数迭代公式和约束如下:

$$b(x,i) = \frac{\sum_{t=1}^T K^{(2)}(t) (U_i(x,t) - B(x)K^{(1)}(t) - \gamma(x,i)g(t,i))}{\sum_{t=1}^T K^{(2)}(t)^2}, \quad (12)$$

$$r(x,i) = \frac{\sum_{t=1}^T g(t,i) (U_i(x,t) - B(x)K^{(1)}(t) - b(x,i)K^{(2)}(t))}{\sum_{t=1}^T g(t,i)^2}, \quad (13)$$

$$\sum_x b(x,i) = 1. \quad (14)$$

3. 预测结果分析

3.1. 数据来源

本文搜寻人类死亡数据库, 选取了中国、智利、保加利亚这 3 个发展中国家和美国、日本、英国这 3 个发达国家的死亡率为研究对象。其中, 智利是南美洲最发达的国家, 保加利亚位于欧洲, 2021 年保加利亚人均 GDP 为 1.22 万美元, 与中国的 1.26 万美元相近。

本文采用的中国死亡率数据来源于 1995~2020 年的《中国人口统计年鉴》, 其他国家的数据均来自人类死亡率数据库, 由于保加利亚数据更新较快, 其研究数据区间为 1995~2021 年, 其余国家死亡率数据均只更新到了 2020 年。由于死亡率数据在高年龄段的瞒报、错报现象较为严重, 故本文只采用了各国各年 0~89 岁的死亡率数据。

3.2. 参数估计结果

根据 2.2 给出的迭代公式, 应用 R 软件对六国 1995~2017 年的 0~89 岁的历史死亡率数据进行参数拟

合, 得到了各全局参数和局部参数的拟合值, 并做出图 3、图 4 和图 5 使其可视化。

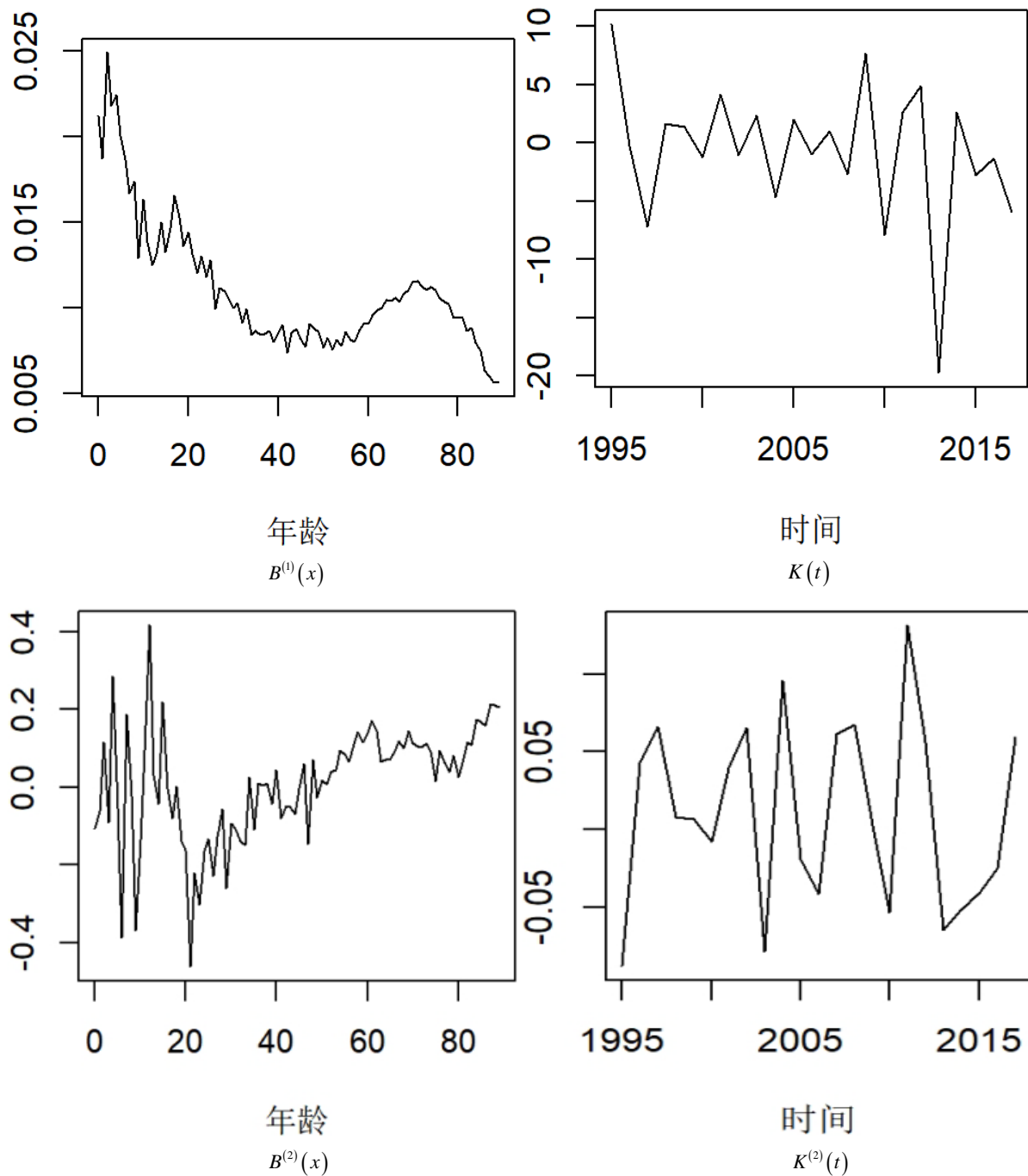


Figure 3. Results of global parameter estimation for ACF1-GDP and ACF3-GDP

图 3. ACF1-GDP 和 ACF3-GDP 全局参数估计结果

在图 3 中, 由于全局参数均由 $U(x, t)$ 通过奇异值分解得出, 所以 $B^{(1)}(x)$, $K(t)$ 也分别为 ACF3-GDP 中的 $B(x)$, $K^{(1)}(t)$ 。

$K(t)$ 变化波动较大, 总体呈较明显的下降趋势, 全局时间参数 $K(t)$ 相较于 $K^{(2)}(t)$ 更能代表全局的共同时间因子, 反映六个国家人口死亡率整体改善状况。

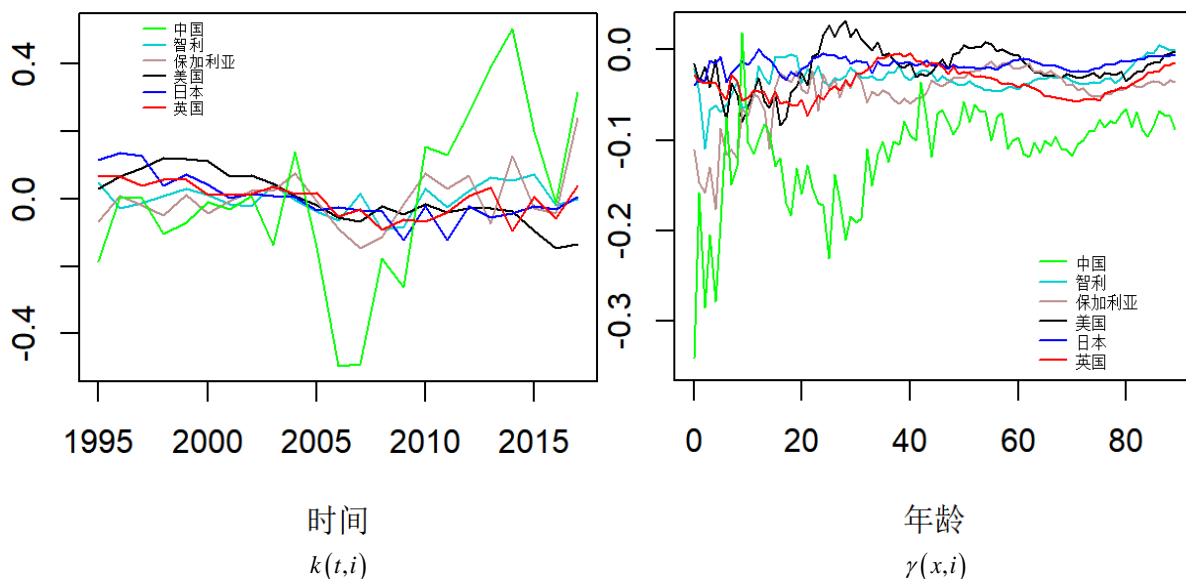


Figure 4. ACF1-GDP local parameter estimation results
图 4. ACF1-GDP 局部参数估计结果

虽然图 3 显示 $B^{(2)}(x)$ 波动幅度较大, 但图 4 中局部年龄 $k(t,i)$ 的值在 0 值附近上下波动, 所以 $B^{(2)}(x)$ 对死亡率的影响较小, $B^{(1)}(x)$ 虽然为正值, 但 $K(t)$ 在大多数年份为负值, 且绝对值较大, 所以使得死亡率下降, 0~20 岁的 $B^{(1)}(x)$ 绝对值要高于其他年龄段, 表示这一年龄段人口死亡率改善状况要优于其他年龄段, 死亡率下降速度更快。同样, 0~20 岁左右以及 80 岁以上的 $B^{(2)}(x)$ 绝对值较大, 表示这两个年龄段人口死亡率改善状况相对优于其他年龄段。

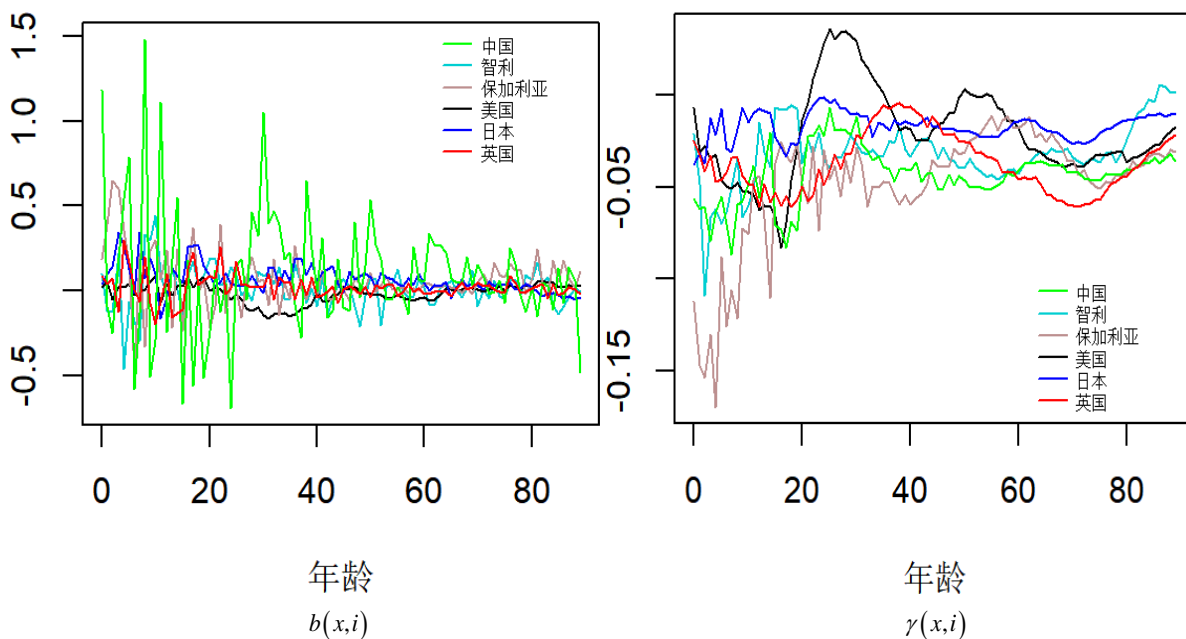


Figure 5. ACF3-GDP local parameter estimation results
图 5. ACF3-GDP 局部参数估计结果

观察对比图 4 ACF1-GDP 模型和图 5 ACF3-GDP 模型中的 $\gamma(x,i)$ 值, 发现只有中国的 $\gamma(x,i)$ 值在两

个模型间具有明显差异,其原因主要在于中国的 $k(t,i)$ 值和 $b(x,i)$ 值相较于其他国家来说波动较大。图4和图5都显示美国的 $\gamma(x,i)$ 值在30岁左右的部分年龄段大于0,这意味着随着经济增长,该年龄段死亡率上升,这代表美国并不适合提出的新模型。

图4和图5中,智利、保加利亚、日本、英国的 $\gamma(x,i)$ 值均为负值,对于这些国家来说,经济增长对各年龄段人口死亡率的降低有积极作用,也更适合本文提出的新模型。且中国、智利、保加利亚0-5岁左右的 $\gamma(x,i)$ 绝对值较高,这说明对于发展中国家,该年龄段人口死亡率下降仍明显受益于经济增长,并且主要受益于随着经济增长带来的医疗水平的提高。而美国的 $\gamma(x,i)$ 值在大部分年龄段围绕0值小幅波动,反映美国人口死亡率下降对经济增长趋势并不敏感,这也部分解释了近年来一些发达国家人口死亡率下降趋势并不明显的事实。

3.3. 预测效果的比较

根据1995~2017年的历史数据对模型的参数进行拟合,然后用带漂移项的随机游走对这23年的时间项参数 $K(t)$, $k(t,i)$ (ACF1、ACF1-GDP中)和 $K^{(1)}(t)$, $K^{(2)}(t)$ (ACF3、ACF3-GDP中)进行预测,这

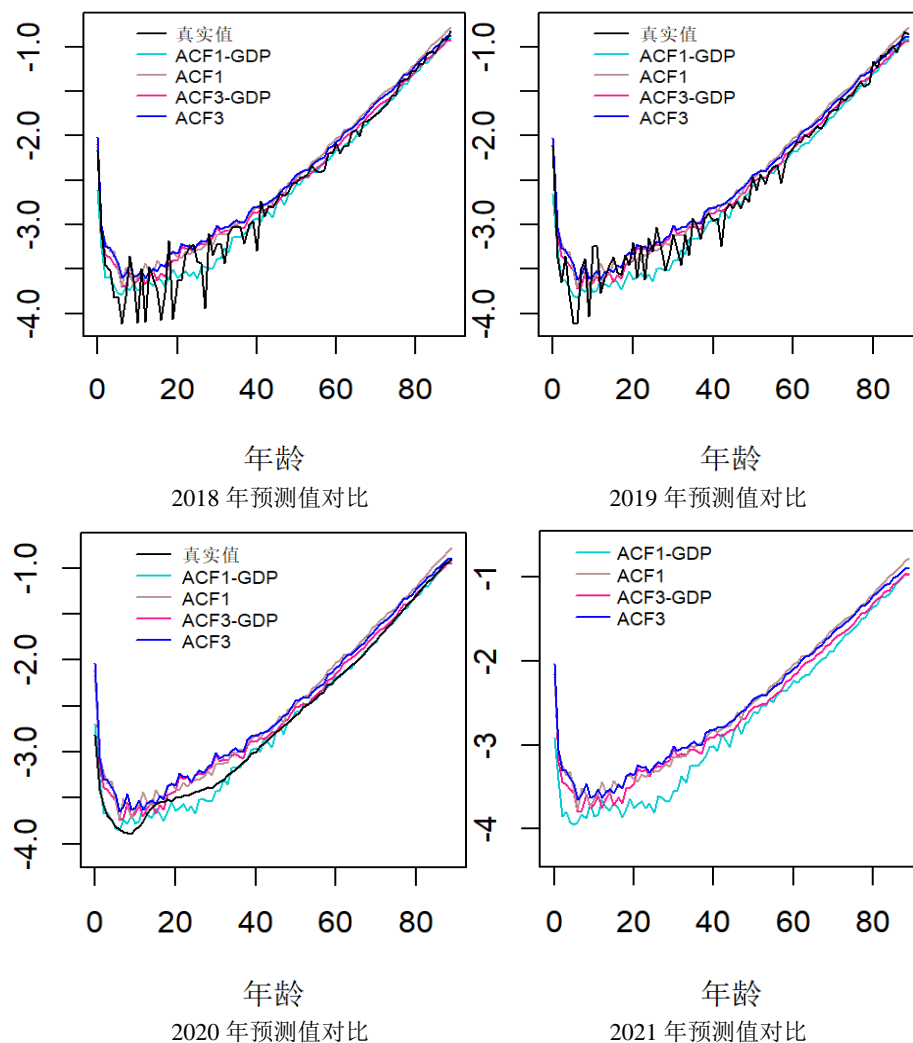


Figure 6. Comparison of logarithmic mortality projections for China, 2018 to 2020

图6. 2018~2020年中国对数死亡率预测值对比

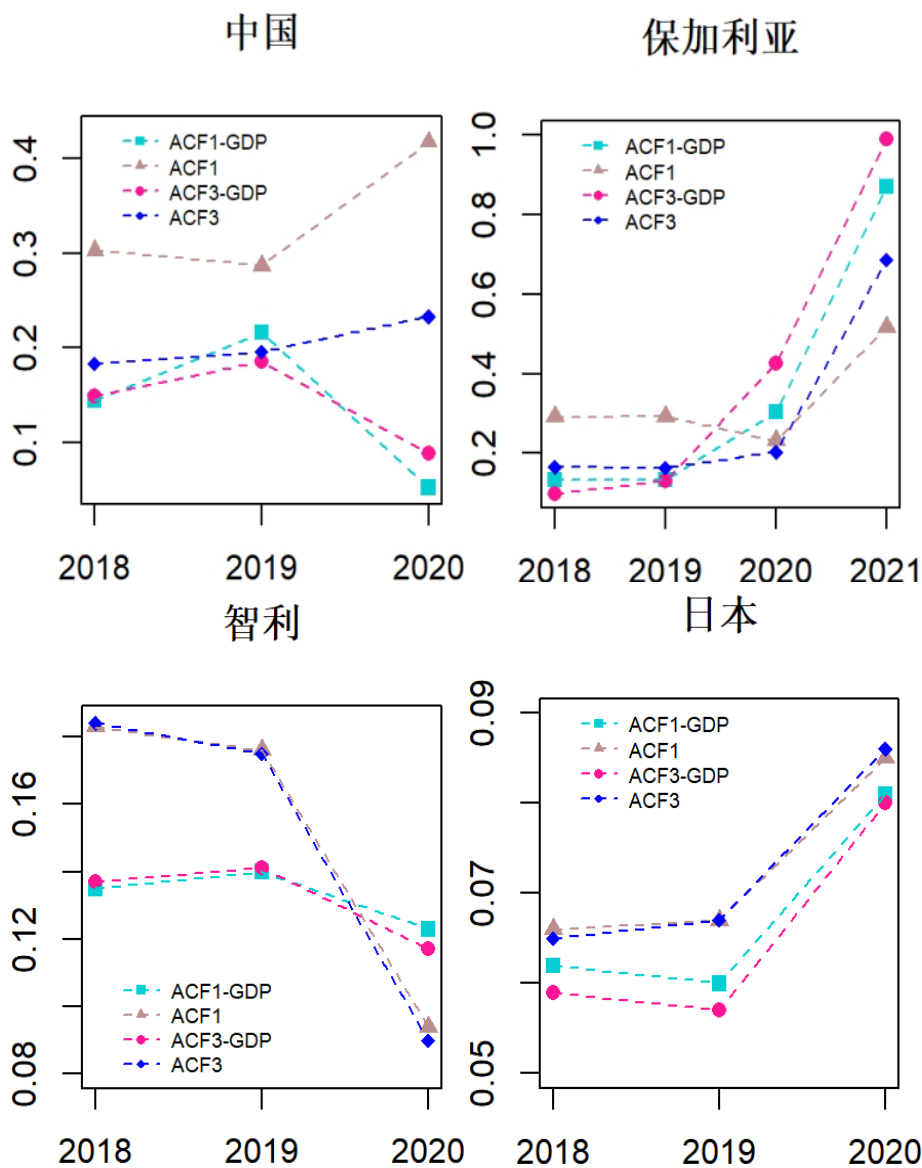
里向后预测了 4 年的时间项, 即从 2018 年到 2021 年, 然后将预测的时间项参数以及拟合得到的各年龄项参数和 $\gamma(x, i)$ 分别代入模型 ACF1-GDP, ACF1, ACF3-GDP, ACF3 中, 得到这 4 年的死亡率预测值。

根据得到的中国死亡率预测值, 做出了 2018~2021 年的对数死亡率预测值对比图(图 6), 可以看出本文提出的新模型 ACF1-GDP 对比 ACF1, ACF3-GDP 对比 ACF3 预测效果较好, 2021 年中国实际的死亡率数据没有更新, 所以只作出了 2021 年四个模型的预测对比。

接下来采用 AFE (绝对拟合误差)来评估 ACF1-GDP, ACF1, ACF3-GDP, ACF3 这四个模型对所研究的 6 个国家的预测效果。AFE 表达式为:

$$AFE_{t,i} = \sum_x |m_{x,t,i} - \hat{m}_{x,t,i}|. \tag{15}$$

其中, 保加利亚的死亡率数据更新到 2021 年, 其他 5 个国家都是更新到了 2020 年。根据这 4 个模型对每个国家每年的死亡率预测的 AFE 值, 我们做出了下面的图:



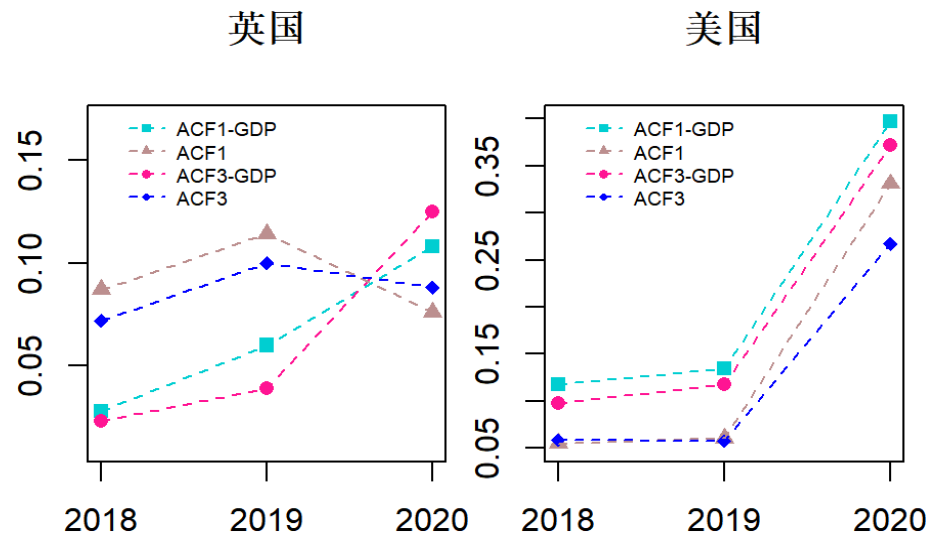


Figure 7. Comparison of errors in mortality projections for six countries
图 7. 六国死亡率预测误差对比

从图 7 中看出, 在这六个国家中, 中国和日本最适合本文提出的加入经济因素的新模型, 因为对于这两个国家, ACF1-GDP 和 ACF3-GDP 的 AFE 值一直都分别比 ACF1, ACF3 的小, 这说明加入经济项的新模型提高了共同因子模型的预测性能。对于保加利亚、智利和英国, 在 2018 年、2019 年 ACF1-GDP 和 ACF3-GDP 模型效果更好, 2020、2021 年时, ACF1 和 ACF3 的 AFE 值更小, 模型效果更好, 分析其原因认为在 2020 和 2021 年, 新型冠状病毒肺炎肆虐, 对欧洲国家的影响较大, 与经济因素相比, 该呼吸道传染病已成为对死亡率影响更大的因素。而对于美国, 从 20 世纪起就一直位于发达国家前列, 经济增长不够明显, 故一直都是不加经济因素的 ACF1, ACF3 模型效果更好, 所以对于类似美国的超发达国家, 本文提出新模型的适用性仍然有待研究。

4. 结论

本文在前人研究基础上, 提出了新模型 ACF1-GDP 和 ACF3-GDP, 使用三个发展中国家(中国、保加利亚、智利)和三个发达国家(美国、日本、英国)的部分历史死亡率数据, 将 ACF1-GDP 模型与 ACF1 模型的预测结果对比, ACF3-GDP 模型与 ACF3 模型预测结果对比, 以此探究加入经济项后的效果是否良好, 结果表明, 在这六个国家中, 中国和日本最适合本文提出的加入经济因素的新模型。

对于保加利亚、智利和英国, 在 2018 年和 2019 年, ACF1-GDP 与 ACF3-GDP 模型效果更好, 在 2020 和 2021 年, ACF1 与 ACF3 的模型效果更好, 分析其原因认为在 2020 和 2021 年, 新型冠状病毒肺炎肆虐, 对欧洲国家的影响较大[11], 与经济因素相比, 该呼吸道传染病已成为对死亡率影响更大的因素。故后续研究可以从加入新型冠状病毒肺炎的影响因素项入手, 为多人口死亡率研究做出新的尝试。在疫情逐渐趋于平稳化的今天, 本文提出的 ACF1-GDP 与 ACF3-GDP 模型对于大部分发达国家与发展中国家的死亡率预测将会有更大的优势。

基金项目

本文由国家自然科学基金项目(12061066)支持。

参考文献

- [1] Li, N. and Lee, R. (2005) Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter

- Method. *Demography*, **42**, 575-594. <https://doi.org/10.1353/dem.2005.0021>
- [2] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011) Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics & Economics*, **48**, 355-367. <https://doi.org/10.1016/j.insmatheco.2010.12.005>
- [3] Brenner, M.H. (2005) Commentary: Economic Growth Is the Basis of Mortality Rate Decline in the 20th Century—Experience of the United States 1901-2000. *International Journal of Epidemiology*, **34**, 1214-1221. <https://doi.org/10.1093/ije/dyi146>
<https://academic.oup.com/ije/article-pdf/34/6/1214/18479307/dyi146.pdf>
- [4] Hanewald, K. (2011) *Explaining Mortality Dynamics: The Role of Macroeconomic Fluctuations and Cause of Death Trends*. *North American Actuarial Journal*, **15**, 290-314. <https://doi.org/10.1080/10920277.2011.10597622>
- [5] Niu, G. and Melenberg, B. (2014) Trends in Mortality Decrease and Economic Growth. *Demography*, **51**, 1755-1773. <https://doi.org/10.1007/s13524-014-0328-3>
<https://read.dukeupress.edu/demography/article/51/5/1755/169414>
- [6] Boonen, T.J. and Li, H. (2017) Modeling and Forecasting Mortality with Economic Growth: A Multipopulation Approach. *Demography*, **54**, 1921-1946. <https://doi.org/10.1007/s13524-017-0610-2>
- [7] 王梦珂, 唐爽. 考虑经济因素的多总体随机死亡率模型改进[J]. *统计与决策*, 2022, 38(5): 37-42. <http://doi.org/10.13546/j.cnki.tjyj.2022.05.007>
- [8] Enchev, V., Kleinow, T. and Cairns, A.J.G. (2017) Multi-Population Mortality Models: Fitting, Forecasting and Comparisons. *Scandinavian Actuarial Journal*, **2017**, 319-342. <https://doi.org/10.1080/03461238.2015.1133450>
- [9] 马海飞. 共同因子死亡率模型及其比较研究[D]: [硕士学位论文]. 兰州: 西北师范大学, 2021.
- [10] Savitzky, A. and Golay, M.J.E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, **36**, 1627-1639. <https://doi.org/10.1021/ac60214a047>
- [11] Spearman, C.W., Aghemo, A., Valenti, L. and Sonderup, M.W. (2021) COVID-19 and the Liver: A 2021 Update. *Liver International*, **41**, 1988-1998. <https://doi.org/10.1111/liv.14984>